both groups, there is no correlation between mastoid pneumatisation and the smallest area of the lumen of the isthmus.

REFERENCES

1. Daniel HJ, Fulghum RS, Brinn JE, Barrett KA. Comparative anatomy of eustachian tube and the middle ear cavity in animal models for otitis media. *Ann Otol Rhinol Laryngol* 1982; **91:** 82-89.
2. Sadé J, Hadas E. Prognostic evaluation of secretory otitis media as a function of mastoid pneumatisation. *Arch Otolaryngol* 1979; **105:** 39-44.
3. Honjo I. Evaluation of eustachian tube function by contrast roentgenography. *Arch Otolaryngol* 1981; **107:** 250-352.
4. Zollner F. *Ohrtrompete.* Berlin: Springer-Verlag, 1942.
5. Bryant WS. Eustachian tube: its anatomy and movement, with a description of the cartilages, muscles, fasciae in the fossa of Rosemuller. *Med Rec NY* 1907; **71:** 931-934.
6. Tomada K, Mori S, Yamoshita T, Kumazawa T. Histology of human eustachian tubal muscles: effect of aging. *Ann Otol Rhinol Laryngol* 1984; **93:** 17-24.
7. Graves GS, Edwards LF. The eustachian tube. *Arch Otolaryngol* 1944; **39:** 359-397.
8. Simkins CS. Functional anatomy of the eustachian tube. *Arch Otolaryngol* 1943; **38:** 476-479.
9. Doyle WS, Rood SB. Comparison of the anatomy of the eustachian tube in the rhesus monkey and man. *Ann Otol Rhinol Laryngol* 1980; **89:** 47-57.
10. Rood SR, Doyel WS. Eustachian tube anatomy. *Ann Otol Rhinol Laryngol* 1985; **94:** suppl. 120, 6-8.
11. Misurya VK. Functional anatomy of tensor palati and levator muscles. *Arch Otolaryngol* 1976; **102:** 265-270.
12. Seiff S, Dellor AL. Anatomic relationship between the human levator and tensor veli palatini muscle and the eustachian tube. *Cleft Palate J* 1978; **15:** 329-336.
13. Rose M. Functional anatomy of the tensor palati, its relevance in cleft palate surgery. *Arch Otolaryngol* 1971; **93:** 1-8.

Epidemiological research methods

Part V. Follow-up studies

D. YACH, J. L. BOTHA

Follow-up studies (also called prospective or cohort studies) are used to determine the natural history of disease, to evaluate the role of risk factors in causation or association, to determine the prognosis of patients with existing disease, and to evaluate the role of drugs and other interventions in preventing disease or further complications. In follow-up studies sampling is prospective, because individuals are followed up to see whether they develop the outcome of interest.[1] These studies have become increasingly important in recent decades with the epidemiological transition that has paralleled population development from a high incidence of acute infectious diseases to a high incidence of chronic, non-infectious diseases in ageing populations.[2]

There are several types of follow-up studies. Randomised controlled trials (RCTs) have been discussed in a previous article.[1] In RCTs the allocation to groups to be followed up and the exposure/intervention being evaluated is under the control of the investigator. In non-RCT follow-up studies the groups are self-selected or defined by nature.

In follow-up studies, baseline measurements on characteristics (risk factors, drugs, disease stage) are made on a sample of the population. On the basis of these measurements subgroups are defined for comparison later. The whole sample is then followed up to estimate the incidence rates or risk of developing the outcome of interest in the subgroups defined at the start.[2]

In this article we will discuss the assembly and follow-up of groups, the assessment of the outcome, and methods of analysis. We illustrate important points by referring mainly to two studies. The first study, conducted by the International Centre for Diarrhoeal Disease Research, investigated the relationship between nutritional status and diarrhoeal incidence and duration in children 2 - 48 months of age in two rural villages in Bangladesh.[3,4] The second study examined the effect of smoking on subsequent mortality in a group of men in the USA.[5] Both these studies involve evaluating the role of risk factors.

Institute for Biostatistics of the South African Medical Research Council and Department of Community Health, University of Stellenbosch, Parowvallei, CP
D. YACH, M.B. CH.B., B.SC. HONS (EPID.), M.P.H.
J. L. BOTHA, M.B. CH.B., M.SC. (CLIN. EPID.)

Defining the group to be followed up

The group to be followed up is often referred to as the cohort, but this word has two meanings.[6] Originally Andvord used it to describe only a group with a common period of birth.[7] His analysis, using this definition in a tuberculosis study, was found useful because '... every generation has its own distinct curve in which the degree of infection in the first years of life becomes an indicator for the mortality value of age groups that follow'. Degree of infection is one example of an experience shared by a group born within a similar time period. We will use cohort in another, broader sense which has become accepted in epidemiology: a group with some common attribute or experience.[8] This may refer to a year (or years) of birth, date of entry into a workforce, or residence in a particular area. In studies of prognosis the selected group is referred to as the 'inception cohort' (identified at an early and uniform point in the course of their disease[9]).

A cohort is often a subset of a larger target population. In such cases the target population and the sampling method should be specified in order to demonstrate clearly whether the sample is unbiased. Often, however, such representativeness is of secondary importance to clear specification of the inception of the cohort, and is not necessary to obtain valid results. This can be demonstrated by our two examples.

In the Bangladesh study 177 children or 94% of the available number in the specified villages were selected for follow-up (20 others were added later). The cohort is therefore representative of those villages but may not be typical of other villages.

In the study on smoking 22 000 American Cancer Society volunteers in 10 states were asked to identify 5 - 10 men with whom they were likely to remain in touch for the next 10 years. In this way 190 134 men were identified. They may not be representative of all men in their areas but may be more representative of friends of volunteers (an already highly select group), and are not representative of women. For valid results the cohorts in both studies had to be free of the outcome of interest: no diarrhoea or lung cancer should be present at the start. For comparisons within cohorts, subgroups should be comparable with regard to inception.

Baseline measurements used in the Bangladesh study involved standard estimates of nutritional status (length, weight and age). In the other study smoking status was determined with the use of a self-administered questionnaire. Men were categorised as never having smoked, being occasional smokers or being regular smokers. In studies investigating the natural history of disease or prognosis, the stage of disease would be the baseline measurement of importance. Cohort subgroups for comparison, formed on the basis of baseline measurements, should be comparable in all respects except for the characteristic being investigated. If comparability is not possible in the design phase, statistical methods of adjustment can be used during analysis, provided information about confounding factors has been collected. In the Bangladesh study the effect of potential confounders of the nutrition-diarrhoea relationship being investigated, such as water and sanitation services, was reduced by using villages with similar services. In the other study, the role of potential confounders of the smoking-mortality relationship, such as exposure to asbestos, was not investigated.

In both the diarrhoea and the smoking-habit studies, control groups were defined from the same population as the study groups. The controls for the study on smoking were those who had 'never smoked', while in the Bangladesh study children with the best nutritional status constituted controls. In many follow-up studies, however, control groups are selected from the general population, which makes the potential for selection bias a major problem. This is particularly the case in occupational studies where the so-called 'healthy worker effect' (one example of selection bias) can distort the interpretation of morbidity and mortality data.[10-13] The healthy worker effect refers to the health status of the employed being better than that of the general population (the latter includes unemployed, sick, disabled and institutionalised people) and tends to mask real occupational health effects, especially those related to chronic diseases.

Follow-up

In follow-up studies, the cohort is usually defined or selected at the start and followed for a predetermined period of time (the follow-up period is concurrent with the planning and execution of the study). The Bangladesh diarrhoea study[5] is an example of this design (sometimes called a concurrent follow-up study[14]). In contrast, follow-up may start at some point in the past, if information were available from a past census, or if a listing of a selected group (factory workers for example) were available; this is referred to as a non-concurrent (historical) prospective study.[14] The renal transplant study reported by O'Donnell *et al.*[15] is an example of this design.

The major source of bias in a prospective study involves losses from the original cohort during the follow-up period.[16] Such losses affect the internal validity of a prospective study. Loss to follow-up is more important than representativeness. The design of a prospective study needs to take account of possible losses and take steps at the recruitment stage to minimise potential attrition over time. Several approaches have been used successfully to reduce losses. In a study of tuberculosis[17] in 25 752 student nurses, nurses were asked at entry into the study to give addresses of 3 relatives or friends with whom they were likely to remain in contact after completing their studies. During the 3 - 5-year follow-up period, 94,76% were traced using these sources. Nurses difficult to contact did not differ from those easily traced, with respect to their risk of tuberculosis or overall mortality. The authors therefore recommended that considerable cost savings would accrue if a sample of non-responders was traced.[17]

In a population-based cardiovascular disease study non-responders differed from responders both in terms of morbidity and in terms of prevalence of risk factors[18] as determined during the baseline study. Since respondents can differ from non-respondents it is important in population-based studies to obtain community support from recognised leaders, the co-operation of local health workers and support of the media at an early stage (and ensure high visibility during the follow-up period) so as to minimise losses.[19] In the Bangladesh study,[4] a health visitor went to each child's home daily to ensure that no children were lost to follow-up. In the smoking-habit study,[5] the volunteers selected only men with whom they were likely to remain in contact; 98,8% of the selected sample were successfully traced during the 20-month follow-up period.

In the RSA few community-based prospective studies have been undertaken. In the Coronary Risk Factor Intervention Study (CORIS) study 7 188 people were included in the inception cohort assembled in 1979. By follow-up in 1983, only 56,7% of the original cohort remained. Those lost to follow-up (i.e. individuals present only in the baseline study) differed from those remaining in terms of several demographic and other characteristics measured in 1979 (P. Jooste — personal communication). The effect on the comparison between subgroups of those lost to follow-up depends not only on the percentage lost, but particularly on the characteristics of the latter.

Outcome assessment

In follow-up studies, as in all other studies, the outcome must be defined in a clear and unambiguous manner. In the Bangladesh study the definition of outcome used was complex. Diarrhoea was defined as being present if a child had more than 4 liquid stools for 1 day. The diarrhoeal episode was considered complete on the first day that the child had fewer than 3 liquid stools. During the study both the health visitor and the doctor used several methods to validate information about diarrhoeal status.[4] For the study on smoking[5] the major outcomes were clearly defined: dead, alive and status unknown.

In a renal transplant study,[15] the outcome of renal transplant patients with impaired renal function at 6 months was assessed. The definition of the major outcome of interest, failed or rejected graft, was not related to an easily validated outcome measure. More objective outcomes that would have been preferable include nephrectomy or dialysis, as recommended recently.[20] In this study it was particularly important that an unambiguous outcome be used since the study spanned a 13-year period during which the criteria used for 'failure' or 'rejection' may have changed.

Analysis of follow-up studies

In follow-up studies an inception cohort is followed for a period of time. After the period, the cohort is examined to determine which participants achieved a predetermined end-point. The end-point may be death, successful response to treatment, relapse or development of a disease. Analysis of

such studies is complicated by two problems, namely the fact that individuals are first observed at different stages of disease (and after different durations of disease), and that some individuals are lost to follow-up.

There are several ways to deal with the first problem. These include prognostic stratification and regression techniques. Prognostic stratification was used in the renal transplant study.[15] Patients were stratified into 3 groups according to their renal function 6 months after the transplant (the start of the study) since the authors' prior belief was that renal function is a prognostic factor in subsequent survival. Regression techniques can be used for the same purpose.[21]

In recent years several techniques for dealing with problems of both staggered entry and variable follow-up have been devised.[21-23] They fall under the general heading of survival analysis. Two basic approaches to the analysis of prospective studies that take into account losses to follow-up involve either calculating an incidence rate or determining the risk of an outcome.

Calculating an incidence rate

In the medical literature prognosis is frequently calculated in terms of the percentage surviving 5 years after diagnosis. The number of those known to be alive at the end of the 5-year period is divided by the number of cases diagnosed (the inception cohort). This approach usually underestimates survival.[22]

The problem of defining the numerator and denominator of a rate can be illustrated by a hypothetical example. Suppose a follow-up study were conducted to determine the annual death rate in workers in a plant between 1 January 1976 and 31 December 1985. If, in calculating the 10-year survival rate the denominator were taken to be all those alive and working on 1 January 1976 (5 workers) and the numerator those alive and working 31 December 1985 (2 workers), the result would be a 10-year survival rate of 40% or a 10-year mortality rate of 60%. The annual death rate could be inferred to be 6% (60/10).

This approach is clearly incorrect if the information in Fig. 1 is used. The stylised figure contains the experience of 11 workers (*a* - *k*). The survival rate above was estimated as 2/5, where *a* and *j* form the numerator and *a, b, c, d,* and *k* the denominator. Worker *j* was not in the original cohort. Persons *b, c, g, h* and *k* were lost to follow-up over the 10-year period. The above calculation assumes that they all died during the interval. Two workers who died in the interval, *e* and *f*, were not included in this calculation. Note that worker *i* was correctly ignored.

A more exact (and exacting) method involves calculating person-time (mentioned in an earlier article[3]). Each person contributes the amount of time they were under observation to the denominator of the rate. For example, person *a* contributed the full 10 years, while person *c* contributed 8 years before being lost to follow-up and person *e* only entered the study in 1976 and died 8 years later. In this way the total person-years under observation can be calculated. In our example, this amounts to a total of 73 person-years. The numerator consists of new events (deaths) reported in the interval and the annual mortality rate is equal to 3/73 or 4,1%.

The assumptions of this method are, first, that those lost to follow-up have the same survival as those under observation and, secondly, that events (deaths or any other outcome) occur uniformly across the time interval.[21] The latter assumption is particularly important when long time intervals are used. The likelihood of death in children, for example, is much higher at the start of the 0 - 5-year age group. When this occurs, shorter time intervals can be used.

A further problem not specific to the analysis of follow-up studies relates to changes that occur over calendar time. In the

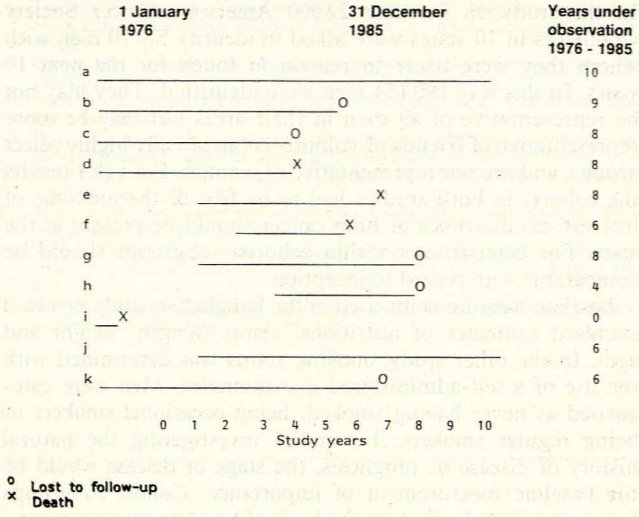

Fig. 1. Variable follow-up in a prospective study.

renal transplant study, for example, patients were entered into the study over a 13-year period during which diagnostic and therapeutic changes that could affect survival are likely to have occurred. In such a situation, separate analyses for calendar periods should be undertaken or more powerful techniques such as Cox's proportional hazards regression methods are needed.[21]

In Table I the results of the Bangladesh study are presented using person-time. In the analysis only days for which reliable information was available were used. The inception cohort was defined at the outset in terms of their nutritional status. Children with a weight-for-length ratio of $\geqslant$ 90%, 80 - 89% and $\leqslant$ 79% were compared with respect to their diarrhoeal incidence and duration. Incidence of diarrhoea was calculated by dividing diarrhoeal episodes by days under observation. The results show that nutritional status was related to (or a risk factor for) diarrhoeal duration but not incidence.

TABLE I. DIARRHOEAL INCIDENCE AND DURATION BY NUTRITIONAL STATUS*

Weight-for-length ratio	Days of observation	Incidence rate (per day)	Duration (days)
$\geqslant$ 90%	8 385	16,9	6,8 $\pm$ 0,9
80 - 89%	13 860	16,2	8,5 $\pm$ 0,8
$\leqslant$ 79%	4 986	16,4	10,6 $\pm$ 1,7

*Similar results were obtained using a weight-for-age ratio.

The incidence rate is used as measure of the force of morbidity or mortality on a population and has no interpretation for an individual. It is used in studies measuring the impact of disease in different groups, as in the Bangladesh study or to assess causality, as in the study in which mortality rates by age/smoking category were computed and compared.

Calculating risk

In many instances it is more important to be able to determine the risk or probability of an individual developing a disease (or changing status) over a certain time period. When one is making clinical or personal decisions about an individual's prognosis or the effect of a particular treatment or behaviour (e.g. smoking), knowledge concerning the risk of

disease for the individual is useful.[24] The life-table approach is used to estimate the risk of death for groups of people over several time intervals. Life-table techniques are also useful for comparing groups with respect to survival at different points over time. Statistical methods exist to calculate confidence intervals for such survival curves.[22] In small studies, the time intervals can be shortened to include only people with the same survival time (the Kaplan-Meier approach[22]). The same assumptions apply to the life-table and Kaplan-Meier approaches as to the person-time approach mentioned earlier.

Follow-up v. case-control studies

The major cost and time savings of case-control studies need to be balanced against the advantages of follow-up studies. In general, case-control studies may be preferable in situations where existing and past information is reliable, where exposure has not changed over time or where a rare disease with a large number of possible causes is being studied. Follow-up studies are advisable in determining the natural history of a disease, or the effect of variable intensities of rare exposures on outcomes. Neither design, however, is free of selection or measurement bias which can distort the results.

REFERENCES

1. Botha JL, Yach D. Epidemiological research methods: III. Randomised controlled trials (for interventions) *S Afr Med J* 1987; **71:** 657-660.
2. Susser M. Epidemiology in the United States after World War II: the evaluation of technique. *Epidemiol Rev* 1985; **7:** 147-177.
3. Black RE, Brown KH, Becker S, Yunus M. Longitudinal studies of infectious diseases and physical growth of children in rural Bangladesh: I. Patterns of morbidity. *Am J Epidemiol* 1982; **115:** 305-314.
4. Black RE, Brown KH, Becker S. Malnutrition is a determining factor in diarrhoeal duration but not incidence among young children in a longitudinal study in rural Bangladesh. *Am J Clin Nutr* 1984; **39:** 87-94.
5. Hammond EC, Horn D. The relationship between human smoking habits and death rates. *JAMA* 1954; **155:** 1316-1328.
6. Last JM. *A Dictionary of Epidemiology*. New York: Oxford University Press, 1983.
7. Sartwell PE. Cohorts: the debasement of a word. *Am J Epidemiol* 1976; **103:** 635-638.
8. Frost WH. The age selection of mortality from tuberculosis in successive decades. *Am J Hygiene* 1939; **30:** 91-96.
9. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston: Little, Brown, 1985.
10. McMichael AJ. Standardized mortality ratios and the 'healthy worker effect': scratching beneath the surface. *J Occup Med* 1976; **18:** 165-168.
11. Fox AJ, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. *Br J Prev Soc Med* 1976; **30:** 225-230.
12. Wen CP, Tsai SP, Gibson RL. Anatomy of the healthy worker effect: a critical review. *J Occup Med* 1983; **25:** 283-289.
13. Sterling TD, Weinkam JJ. The 'healthy worker effect' on mortality rates. *J Occup Med* 1985; **27:** 477-482.
14. Lilienfield AM, Lilienfield DE. *Foundations of Epidemiology*. New York: Oxford University Press, 1980: 226-255.
15. O'Donnell D, Seggie JL, Dawes P, Botha JR, Meyers AM, Myburgh JA. Outcome of renal transplants with impaired renal function at 6 months. *S Afr Med J* 1987; **71:** 71-72.
16. Kleinbaum DG, Morgenstern H, Kupper LL. Selection bias in epidemiologic studies. *Am J Epidemiol* 1981; **113:** 452-463.
17. Theodore A, Berger AG, Palmer CE. A follow-up study of tuberculosis in former student nurses: I. Methods of locating and obtaining information from a study population of 25 752. *J Chron Dis* 1956; **3:** 499-520.
18. Criqui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. *Am J Epidemiol* 1978; **108:** 367-372.
19. Croft JB, Webber LS, Parker FC, Berenson GS. Recruitment and participation of children in a long-term study of cardiovascular disease: the Bogalusa heart study, 1973 - 1982. *Am J Epidemiol* 1984; **120:** 436-448.
20. The Canadian Multicentre Transplant Study Group. A randomized clinical trial of cyclosporine in cadaveric renal transplantation: analysis at 3 years. *N Engl J Med* 1986; **314:** 1219-1225.
21. Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman and Hall, 1984.
22. Lee ET. *Statistical Methods for Survival Data Analysis*. Belmont, Calif.: Wadsworth, 1980.
23. Kahn HA. *An Introduction to Epidemiologic Methods*. New York: Oxford University Press, 1983.
24. Morgenstern H, Kleinbaum DG, Kupper LL. Measures of disease incidence used in epidemiologic research. *Int J Epidemiol* 1980; **9:** 97-104.

Nuus en Kommentaar/News and Comment

Malnutrition in the hospital

Here we go again. Every time someone investigates the nutritional status of patients in a hospital horrible things are discovered. Why is it that the patient who is receiving the best of medical and nursing care in a centre of excellence is not receiving an adequate diet?

The latest report comes from a team of Swedish nutrition experts from Göteborg who have been carrying out investigations in Scandinavian hospitals. A couple of years ago they found that in a teaching hospital surgical patients were going short on food, and they now report on their findings in a gastro-enterology department of another major hospital in Norway (Waage *et al.*, *Tidsskr Nor Laegeforen* 1987; **107:** 146). It is true that it wasn't altogether the hospital's fault since of the 26 patients (13 women and 13 men) of average age 60 years, 6 were regarded as undernourished by the usual anthropometric standards on admission. These were all older patients. However, when the in-hospital nutritional intake was recorded, 13 out of 20 (65%) had an energy intake under 80% of requirement and 15 out of the 20 (75%) had a protein intake under 80% of requirement. In fact, further enquiries suggested that 14 patients out of the 20 had a lower energy intake when they were in hospital than they had at home.

Does this matter? The authors think that it does because malnourished patients are more liable to complications. They had demonstrated this already in a surgical ward and see no reason to think that matters were any different in a medical one. The observation period was short, but other reports from America, Sweden and Denmark have demonstrated the same type of situation with longer observation periods.

Recurrences in oral and genital herpes simplex infection

Oropharyngeal herpes is not nearly as troublesome an infection as genital herpes and although there is great variability among individuals in the clinical recurrence rates of both, it seems tht oral herpes does not recur as frequently as genital herpes. Lafferty *et al.* (*N Engl J Med* 1987; **316:** 1444) studied the natural history of primary herpes simplex virus (HSV) infection in 39 adults who had concurrent infection of the oropharynx and the genitalia. Twelve of the patients were infected with HSV type 1 and 27 with HSV type 2, the same virus being responsible for infection at both sites.

Their report shows that subsequent recurrence patterns differed markedly according to the viral type and the anatomical site. The frequency of recurrence after genital HSV-2 infection was higher than that after oral-labial HSV-1 infection, while genital HSV-1 recurrences were more common than oral-labial HSV-2 recurrences. Overall genital recurrenes were nearly 6 times more frequent than oral-labial recurrences during a median 380 days' follow-up. Five of the 12 patients with HSV-1 infections subsequently reported an oral-labial recurrence compared with only 1 of the 27 patients with HSV-2 infections. Conversely 24 of 27 patients from whom HSV-2 was isolated from the genital region, compared with 3 of 12 from whom HSV-1 was isolated, had a recurrence of genital disease.

The findings may account for the relatively rapid increase in prevalence of genital herpes in recent years.